

Establishing χ^2 Properties of
Sums of Squares Using Induction
(and no Matrices)

Shayle R. Searle
Friedrich Pukelsheim

Preprint Nr. 55

Februar 1985

BU-869-M

Establishing χ^2 Properties of Sums of Squares Using Induction (and no Matrices)

Shayle R. Searle * and Friedrich Pukelsheim

**Institut für Mathematik
Universität Augsburg, Federal Republic of Germany**

ABSTRACT

The between-classes sum of squares in a between-and-within classes analysis of variance has, under normality, a χ^2 distribution. Although "substantial mathematical machinery" (Stigler, 1984) is often used in classroom derivation of this distribution, it can be avoided by using induction and independence properties of standard normal variables. This is the derivation given here — for unequal-subclass-numbers data. Independence of the between-and within-classes sums of squares is also shown.

*On leave from Cornell University, Ithaca, N.Y., supported by a U. S. Senior Scientist Award from the Alexander von Humboldt-Stiftung.

1. Introduction

Stigler (1984) points out, quite rightly, that "In introductory courses in mathematical statistics, the proof that the sample mean \bar{X} and sample variance S^2 are independent when one is sampling from normal populations is commonly deferred until substantial mathematical machinery has been developed". Contrasting this, Stigler then gives a very nice proof for the one-sample case that requires understanding nothing more than normality and independence, together with the definition of a χ^2 variable as the sum of squares of independent and identically distributed standard normal variables. The sum of independent χ^2 variables being distributed as χ^2 is also used. His method of proof, which relies on induction on sample size, is extended here to sums of squares in a one-way classification with unbalanced data (unequal subclass numbers data). It is in this situation, of the analysis of variance of unbalanced data, that Stigler's "substantial mathematical machinery" is, generally speaking, nowhere more evident, so that for teaching purposes there are great advantages in being able to avoid such complexities as is done here.

For observations y_{ij} for $i = 1, \dots, a$ and $j = 1, \dots, n_i$ assume that observations having the same value of i (those in the i 'th class) are identically distributed with a normal density having mean μ_i and variance σ^2 . Write this as

$$y_{ij} \sim \text{i.i.d. } N(\mu_i, \sigma^2) \quad \text{for } j = 1, \dots, n_i; \quad (1)$$

and let this be true for each $i = 1, \dots, a$. Also assume that observations in each class are independent of those in every other class. Thus, using $v(y_{ij})$ for the variance of y_{ij} and $\text{cov}(y_{ij}, y_{hk})$ for the covariance between y_{ij} and y_{hk} ,

$$v(y_{ij}) = \sigma^2 \quad \text{and} \quad \text{cov}(y_{ij}, y_{hk}) = 0 \quad (2)$$

for all i, j and for all h, k except $i = h$ with $j = k$.

The sample mean of the observations in class i will be denoted by

$$\bar{y}_i = \sum_{j=1}^{n_i} y_{ij} / n_i. \quad (3)$$

We define (for subsequent convenience) partial sums of the n_i -values:

$$s_i = \sum_{t=1}^i n_t, \text{ and particularly } s_a = \sum_{t=1}^a n_t = n_{..} \quad (4)$$

And the mean of all observations in all a classes will be denoted by

$$m_a = \bar{y}_{..} = \sum_{i=1}^a n_i \bar{y}_i / \sum_{i=1}^a n_i = \sum_{i=1}^a n_i \bar{y}_i / s_a = \sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij} / s_a. \quad (5)$$

Then from (2), (3) and (5)

$$v(\bar{y}_i) = \sigma^2 / n_i \quad \text{and} \quad v(m_a) = \sigma^2 / s_a. \quad (6)$$

$$\text{cov}(y_{ij}, \bar{y}_i) = \sigma^2 \quad \text{and, for } i \neq i', \quad \text{cov}(\bar{y}_i, \bar{y}_{i'}) = 0; \quad (7)$$

and, for $i = 1, \dots, a$

$$\text{cov}(\bar{y}_i, m_a) = n_i \sigma^2 / (n_i s_a) = \sigma^2 / s_a. \quad (8)$$

Four distributional results are taken as known: (i) that normal variables having zero covariance are independent; (ii) that linear combinations of normal variables are normally distributed; (iii) that a χ_k^2 variable (having k degrees of freedom) is definable as the sum of squares of k i.i.d. normal standard variables; and (iv) that sums of independent χ^2 variables are χ^2 variables. These results are referred to frequently in what follows.

We deal with between-class and within-class sums of squares defined for a classes as

$$B_a = \sum_{i=1}^a n_i (\bar{y}_i - m_a)^2 = \sum_{i=1}^a n_i \bar{y}_i^2 - s_a m_a^2 \quad (9)$$

and

$$W_a = \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad (10)$$

with

$$W_a = \sum_{i=1}^a W_i \quad \text{for} \quad W_i = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2. \quad (11)$$

2. Independence

The independence of B_a and W_a stems directly from (i). Consider one of the terms in B_a of (9) that is squared, say $\bar{y}_i - m_a$; and a similar term from W_a of (10), say $y_{hj} - \bar{y}_h$. The covariance of these terms for $i = h$ is, from (7) and (8)

$$\begin{aligned} \text{cov}(\bar{y}_i - m_a, y_{ij} - \bar{y}_i) &= \text{cov}(\bar{y}_i, y_{ij}) - v(\bar{y}_i) - \text{cov}(m_a, y_{ij}) + \text{cov}(m_a, \bar{y}_i) \\ &= \sigma^2(1/n_i - 1/n_i - 1/s_a + 1/s_a) \\ &= 0; \end{aligned}$$

and, similarly for $i \neq h$,

$$\text{cov}(\bar{y}_i - m_a, y_{hj} - \bar{y}_h) = \sigma^2(0 - 0 - 1/s_a + 1/s_a) = 0.$$

Thus $\bar{y}_i - m_a$ and $y_{hj} - \bar{y}_h$ have zero covariance and so, since by (ii) each of them is normally distributed, they are by (i) independent. Since this is so for all i, h and j it is true for every pair of terms that are squared, one in B_a and one in W_a . Therefore B_a and W_a are independent.

3. Induction

By induction on a we show that $B_a/\sigma^2 \sim \chi_{a-1}^2$. The starting point is the case of $a = 2$. From (9)

$$\begin{aligned} B_2 &= n_1 \bar{y}_1^2 + n_2 \bar{y}_2^2 - (n_1 \bar{y}_1 + n_2 \bar{y}_2)^2 / (n_1 + n_2) \\ &= n_1 n_2 (\bar{y}_1 - \bar{y}_2)^2 / (n_1 + n_2). \end{aligned}$$

From (6) and (7)

$$v(\bar{y}_1 - \bar{y}_2) = \sigma^2(1/n_1 + 1/n_2) = (n_1 + n_2)\sigma^2/n_1n_2, \quad (12)$$

and so, under the hypothesis $H : \mu_1 = \mu_2$, (12) and (ii) yield

$$\bar{y}_1 - \bar{y}_2 \sim N[0, (n_1 + n_2)\sigma^2/n_1n_2].$$

Therefore on defining $g = \sqrt{n_1n_2/(n_1 + n_2)}(\bar{y}_1 - \bar{y}_2)$, we have $g \sim N(0, 1)$, and so, since $B_2/\sigma^2 = g^2$, (iii) gives $B_2/\sigma^2 \sim \chi_1^2 = \chi_{2-1}^2$. Thus $B_a/\sigma^2 \sim \chi_{a-1}^2$ is certainly true for $a = 2$. With this as a base we now show that assuming $B_a/\sigma^2 \sim \chi_{a-1}^2$ yields $B_{a+1}/\sigma^2 \sim \chi_a^2$; i.e., induction on a establishes that $B_a/\sigma^2 \sim \chi_a^2$ is true generally.

Relationships between m_a and m_{a+1} , and B_a and B_{a+1} , are needed that are extensions of well-known recurrence formulae for sample means and variances given in Searle (1983) and Stigler (1984). First, from (4) and (5)

$$\begin{aligned} m_{a+1} &= \sum_{i=1}^{a+1} n_i \bar{y}_i / s_{a+1} \\ &= (s_a m_a + n_{a+1} \bar{y}_{a+1}) / (s_a + n_{a+1}) \\ &= m_a + n_{a+1}(\bar{y}_{a+1} - m_a) / s_{a+1}; \end{aligned} \quad (13)$$

and second, from (9)

$$B_{a+1} = \sum_{i=1}^{a+1} n_i \bar{y}_i^2 - s_{a+1} m_{a+1}^2$$

and on using (13) this is

$$\begin{aligned} B_{a+1} &= \sum_{i=1}^a n_i \bar{y}_i^2 + n_{a+1} \bar{y}_{a+1}^2 - s_{a+1} [m_a + n_{a+1}(\bar{y}_{a+1} - m_a) / s_{a+1}]^2 \\ &= \sum_{i=1}^a n_i \bar{y}_i^2 + n_{a+1} \bar{y}_{a+1}^2 - (s_a + n_{a+1}) m_a^2 \\ &\quad - 2m_a n_{a+1}(\bar{y}_{a+1} - m_a) - n_{a+1}^2(\bar{y}_{a+1} - m_a)^2 / s_{a+1} \\ &= \sum_{i=1}^a n_i \bar{y}_i^2 - s_a m_a^2 + n_{a+1}(\bar{y}_{a+1} - m_a)^2 - n_{a+1}^2(\bar{y}_{a+1} - m_a)^2 / s_{a+1} \\ &= B_a + n_{a+1}(1 - n_{a+1}/s_{a+1})(\bar{y}_{a+1} - m_a)^2, \end{aligned}$$

i. e.,

$$B_{a+1} = B_a + \delta \text{ for } \delta = (n_{a+1}s_a/s_{a+1})(\bar{y}_{a+1} - m_a)^2. \quad (14)$$

Now \bar{y}_{a+1} and m_a are independent, and so

$$v(\bar{y}_{a+1} - m_a) = \sigma^2(1/n_{a+1} + 1/s_a) = s_{a+1}\sigma^2/n_{a+1}s_a.$$

Hence, under the hypothesis

$$H : \mu_1 = \mu_2 = \dots = \mu_{a+1} \quad (15)$$

$$\bar{y}_{a+1} - m_a \sim N(0, \frac{s_{a+1}}{n_{a+1}s_a} \sigma^2).$$

Thus, just as in deriving $B_2/\sigma^2 \sim \chi_1^2$, we have $\delta/\sigma^2 \sim \chi_1^2$. Furthermore, $B_a = \sum_{i=1}^a n_i(\bar{y}_i - m_a)^2$ and for $i = 1, \dots, a$, $\text{cov}(\bar{y}_i - m_a, \bar{y}_{a+1} - m_a) = \sigma^2(0 - 1/s_a - 0 + 1/s_a) = 0$. Therefore, since $\bar{y}_i - m_a$ and $\bar{y}_{a+1} - m_a$ are by (ii) normally distributed, they are by (i) independent. Therefore in (14) B_a and δ are independent; and so with $B_a/\sigma^2 \sim \chi_{a-1}^2$ and $\delta/\sigma^2 \sim \chi_1^2$, this independence means from (iv) that $B_{a+1}/\sigma^2 = B_a/\sigma^2 + \delta/\sigma^2 \sim \chi_{a-1+1}^2 = \chi_a^2$. Thus is the χ^2 property of B_a proven, without recourse to any "substantial mathematical machinery".

4. The within-class sum of squares

The χ^2 property of W_a can now be derived from that of B_a . Suppose that $n_i = 1$ for $i = 1, \dots, a$. Then B_a becomes $\sum_{i=1}^a (y_i - \bar{y})^2$ for $\bar{y} = \sum_{i=1}^a y_i/a$, and so by the immediately preceding result, $\sum_{i=1}^a (y_i - \bar{y})^2/\sigma^2 \sim \chi_{a-1}^2$. A special case of this is W_i of (11): $W_i/\sigma^2 = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2/\sigma^2 \sim \chi_{n_i-1}^2$. Hence, since the W_i s are distributed independently, $W/\sigma^2 = \sum_{i=1}^a W_i/\sigma^2$ has, by (iv), a χ^2 distribution on $\sum_{i=1}^a (n_i - 1) = n. - a$ degrees of freedom, i.e., $W/\sigma^2 \sim \chi_{n.-a}^2$.

5. Application

The ultimate application of these results is, of course, that under the hypothesis (15), which gives $B_a/\sigma^2 \sim \chi_{a-1}^2$ independently of $W_a/\sigma^2 \sim \chi_{n.-a}^2$, the ratio

$$F = \frac{(B_a/\sigma^2)/(a-1)}{(W_a/\sigma^2)/(n.-a)} = \frac{B_a/(a-1)}{W_a/(n.-a)}$$

has the F -distribution on $(a-1)$ and $(n.-a)$ degrees of freedom, and can be used as a test-statistic for the hypothesis (15).

6. Extending Helmert's transformation

In the simple case of x_i for $i = 1, \dots, n$ with $x_i \sim \text{i.i.d. } N(0, \sigma^2)$, the χ_{n-1}^2 distribution of $S^2/\sigma^2 = \sum_{i=1}^n (x_i - \bar{x})^2/\sigma^2$ can be derived using what is known (e.g., Lancaster, 1972) as Helmert's transformation:

$$u_i = \sum_{j=1}^i \lambda_{ij} x_j \quad \text{for } i = 2, \dots, n \quad (16)$$

with

$$\lambda_{ij} = 1/\sqrt{i(i-1)} \quad \text{for } j = 1, 2, \dots, i-1,$$

and

$$\lambda_{ii} = -\sqrt{(i-1)/i}.$$

It is then easily shown that the u_i of (16) are i.i.d. $N(0, \sigma^2)$ and that $S^2 = \sum_{i=2}^n u_i^2$; then (iv) gives $S^2/\sigma^2 \sim \chi_{n-1}^2$.

An extension of (16) provides a second proof that $B_a/\sigma^2 \sim \chi^2_{a-1}$. It uses

$$z_i = \sum_{j=1}^i t_{ij} \bar{y}_j \quad \text{for } i = 2, \dots, a \quad (17)$$

with

$$t_{ij} = n_j \sqrt{n_i / s_{i-1} s_i} \quad \text{for } j = 1, \dots, i-1$$

and

$$t_{ii} = -\sqrt{n_i s_{i-1} / s_i}.$$

It can then be shown that $v(z_i) = \sigma^2$ and $\text{cov}(z_i, z_h) = 0$ for all $i \neq h = 2, \dots, a$; and that $\sum_{i=2}^a z_i^2 = B_a$. Hence the z_i are i.i.d. $N(0, \sigma^2)$ and so $B_a/\sigma^2 \sim \chi^2_{a-1}$.

In passing, observe that $n_i = 1$ for all i reduces t_{ij} and t_{ii} of (17) to λ_{ij} and λ_{ii} respectively, of (16) — as one would expect.

A final, matrix comment is not out of order: on defining $t_{ij} = 0$ for $j = i+1, \dots, a$ and $i = 2, \dots, a$, the resulting $(a-1) \times a$ matrix $\mathbf{T} = \{t_{ij}\}$ for $i = 2, \dots, a$ and $j = 1, \dots, a$ is related to a more general Helmert-style matrix of Irwin (1942), quoted as \mathbf{H} in (4) of Lancaster (1972). The relationship is

$$\mathbf{H} = \begin{bmatrix} \mathbf{n}' / \sqrt{s_a} \\ \mathbf{T} \end{bmatrix} \mathbf{D}$$

where \mathbf{n}' is the row vector $[n, \dots, n_a]$ and \mathbf{D} is the diagonal matrix of diagonal elements $1/\sqrt{n_1}, \dots, 1/\sqrt{n_a}$.

References

- Irwin, J. O. (1942): On the distribution of a weighted estimate of variance and on analysis of variance in certain cases of unequal weighting. *Journal of the Royal Statistical Society, Series A*, 105, 115-118.
- Lancaster, H. O. (1972): The Helmert matrices. *The American Mathematical Monthly* 65, 5-12.
- Searle, S. R. (1983): The recurrence formulae for means and variances. *Teaching Statistics* 5, 7-10.
- Stigler, S. M. (1984): Kruskal's proof of the joint distribution of \bar{X} and S^2 . *The American Statistician* 38, 134-135.

Bestellungen nimmt entgegen:

Institut für Mathematik
der Universität Augsburg
Memminger Str. 6
8900 Augsburg

Verzeichnis der erschienenen Preprints

1. Hoffmann, K.-H.; Sprekels, J.: Real-time Control of the Free Boundary in a Two-phase Stefan Problem (1982), Numer. Funct. Anal. and Optimiz. 5, 47-76 (1982)
2. Schert, R.: Eine Bemerkung zu einer Arbeit von C. Meyer über Imaginär-quadratische Körper mit der Klassenzahl 2 (1982)
3. Brokate, M.: A Hyperbolic Free Boundary Problem: Existence, Uniqueness and Discretization (1982)
4. Grötschel, M.; Padberg, M.: Polyhedral Aspects of the Traveling Salesman Problem I: Theory (1983)
5. Grötschel, M.; Padberg, M.: Polyhedral Aspects of the Traveling Salesman Problem II: Computation (1983)
6. Ritter, J.; Sehgal, S. K.: On a Conjecture of Zassenhaus on Torsion Units in Integral Group Rings (1983), Math. Ann. 264, 257-270 (1983)
7. Barahona, F.; Grötschel, M.; Mahjoub, A. R.: Facets of the Bipartite Subgraph Polytope (1983)
8. Grötschel, M.; Jünger, M.; Reinelt, G.: A Cutting Plane Algorithm for the Linear Ordering Problem (1983)
9. Grötschel, M.; Jünger, M.; Reinelt, G.: Optimal Triangulation of Large Real World Input-output Matrices (1983), Statistische Hefte, 25 (1984), 261-295
10. Jünger, M.; Pulleyblank, W. R.; Reinelt, G.: On Partitioning the Edges of Graphs into Connected Subgraphs (1983)
11. Grötschel, M.: On Intersections of Longest Cycles (1983), "Graph Theory and Combinatorics", B. Bollobás, London, Academic Press, 1984, 171-189
12. Hoffmann, K.-H.; Sprekels, J.: Real-time Control in a Free Boundary Problem Connected with the Continuous (1983) Casting of Steel, Intern. Ser. Numer. Math. 68, Birkhäuser-Verlag (1983)

13. Brokate, M.: Numerical Solution of a Hyperbolic Free Boundary Problem with a Method of Characteristics (1983)
14. Hoffmann, K.-H.; Sprekels, J.: Exact Bounds for the Radially Symmetric Shape of Confined Plasma in the Unit Circle (1983), Math. Meth. in Appl. Sci. 6 (1984) 496-511
15. Grötschel, M.: Polyhedral Combinatorics: An Annotated Bibliography (1983)
16. Euler, R.; Jünger, M.; Reinelt, G.: Generalizations of Cliques, Odd Cycles and Anticycles and their Relation to Independence System Polyhedra (1983)
17. Hoffmann, K.-H.; Sprekels, J.: On the Identification of Coefficients of Elliptic Problems by Asymptotic Regularization (1983)
18. Hoffmann, K.-H.; Sprekels, J.: A Note on the Thermostat Control of Heating Processes in Materials with Thermal Memory (1983), Mat. Aplic. Comp., V. 3, n°. 3, pp. 257 a 264, 1984
19. Alt, H. W.; Hoffmann, K.-H.; Sprekels, J.: A Numerical Procedure to Solve Certain Identification Problems (1983), Intern. Ser. Numer. Math. 68, Birkhäuser-Verlag
20. Brüning, J.: Heat Equation Asymptotics for Singular Sturm-Liouville Operators (1983), Math. Ann. 268, 173-196 (1984)
21. Knabner, P.: A Free Boundary Problem Arising from the Leaching of Saline Soils (1983)
22. Knabner, P.: Regularization of the Cauchy Problem for the Heat Equation by Norm Bounds (1984)
23. Brüning, J.; Heintze, E.: Spektrale Starrheit gewisser Drehflächen (1984), Math. Ann. 269, 95-101 (1984)
24. Hoffmann, K.-H.; Sprekels, J.: On the Identification of Parameters in General Variational Inequalities by Asymptotic Regularization (1984)
25. Hoffmann, K.-H.; Kornstaedt, H.-H.; Sprekels, J.: Automatische Steuerung freier Ränder bei Stefan Problemen durch Thermostatkontrollen am festen Rand (1984)

26. Grötschel, M.: Developments in Combinatorial Optimization (1984), "Perspectives in Mathematics" (Anniversary of Oberwolfach 1984), W. Jäger, J. Moser & R. Remmert, Basel, Birkhäuser-Verlag, 1984, 249-294
27. Knabner, P.: Control of Stefan Problems by Means of Linear-Quadratic Defect Minimization (1984)
28. Grötschel, M.; Jünger, M.; Reinelt, G.: Acyclic Subdigraphs and Linear Orderings: Polytopes, Facets and a Cutting Plane Algorithm (1984)
29. Müller-Funk, U.; Pukelsheim, F.; Witting, H.: On the Duality between Locally Optimal Tests and Optimal Experimental Designs (1984)
30. Jürgens, U.: The Estimation of a Nonstationary Random Coefficient AR(1) Process (1984)
31. Brüning, J.: Spectral Analysis of Singular Sturm-Liouville Problems with Operator Coefficients (1984)
32. Schröder, H.: Some Continuous Index Theorems (1984)
33. Zehendorf, E.: A Non-Existence Theorem for Cyclic MDS-Codes (1984)
34. Christof, K.; Pukelsheim, F.: Approximate Design Theory for a Simple Block Design with Random Block Effects (1984)
35. Grötschel, M.; Lovász, L.; Schrijver, A.: Relaxations of Vertex Packing (1984)
36. Barahona, F.; Grötschel, M.: On the Cycle Polytope of a Binary Matroid (1984)
37. Grötschel, M.; Holland, O.: Solving Matching Problems with Linear Programming (1984)
38. Grötschel, M.: Polyedrische Kombinatorik und Schnittebenenverfahren (1984), Bericht Nr. 231 (1984) der Mathematisch-Statistischen Sektion, Forschungszentrum Graz, Österreich
39. Gurgel, M. A.; Wakabayashi, Y.: On k -leaf connected graphs (1984)
40. Hoffmann, K.-H.; Sprekels, J.: The Method of Asymptotic Regularization and Restricted Parameter Identification Problems in Variational Inequalities (1984)

41. Kielhöfer, H.: Multiple Eigenvalue Bifurcation for Fredholm Operators (1984)
42. Baksalary, J. K.; Pukelsheim, F.: A Note on the Matrix Ordering of Special C-Matrices (1984)
43. Pawlow, I.; Niezgódka, M.: Discrete Approximation of Multi-Phase Stefan Problems with Possible Degeneration (1984)
44. Pawlow, I.: Boundary Control of Degenerate Two-Phase Stefan Problems (1984)
45. Kielhöfer, H.: Interaction of Periodic and Stationary Bifurcation from Multiple Eigenvalues (1984)
46. Grötschel, M.: Fortschritte in der polyedrischen Kombinatorik (1984)
47. Schröder, H.: On the Homotopy type of the regular group of a real W^* -algebra (1984)
48. Pukelsheim, F.: A Note on Predictable Criteria for Absolute Continuity and Singularity of Two Probability Measures (1984)
49. Knabner, P.: Global Existence in a General Stefan-like Problem (1984)
50. Hoffmann, K.-H.; Sprekels, J.: On the Identification of Heat Conductivity and Latent Heat in a One-Phase Stefan Problem (1985)
51. Appell, J.: Deux Méthodes Topologiques pour la Résolution des Équations Elliptiques Non linéaires sans Compacité
52. Appell, J.; de Pascale Espedito: Théorèmes de Bornage pour l'Opérateur de Nemyckii dans les Espaces Idéaux
53. Appell, J.: Misura di Non Compattezza in Spazi Ideali